

Søgning via emnehierarki i bibliotek.dk – hvorfor nu det?

Af Kirsten Larsen

Artiklen er baseret på et oplæg på temadagen "Redskabs-gymnastik med emneord og klassifikation eller frit fald i Nettet?" på Vejle Bibliotek 27. januar 2005 arrangeret af DF's Forum for Registrering og giver en baggrund for udviklingen af en ny søgemåde i bibliotek.dk

Af Kirsten Larsen

Afdelingschef - afdelingen for bibliotek.dk og DanBib Netdivisionen Dansk Biblioteks-Center
kl@dbc.dk



Baggrund for ideen om emnehierarki som søgeindgang

Søgemåderne i bibliotek.dk er i dag flere:

- Søgning, Artikelsøg og Musiksøg er alle baseret på en slags "formularsøg" – altså hvor man "hjælper" brugeren til at anføre om søgekriteriet nu er forfatter, titel eller...
- Slå op i register – mulighed for skanning i de tre vigtigste sætningsregistre
- Kommandosøg – kun for inkarnerede ccl-brugere

Dertil kommer diverse muligheder for videresøgning via links i posterne, Søg litteratur om..., Find titler der ligner.

Vi har arbejdet med at hjælpe brugerne af bibliotek.dk til bedre søgeresultater på en række områder – ikke alle er færdige:

- Forbedring af match
- Sortering
- Litteratur om forfatteren
- Noget der ligner
- Simpel rangordning efter emne
- Henvisning til Biblioteksvagten
- Emnehierarki – i arbejde
- Værkvisning – i arbejde
- Best match – på planen for 2005
- Mellemresultater vises – på planen for 2005

Fra bibliotek.dk's start har vi overvejet om der som alternativ til disse traditionelle søgemåder, også skulle være en "Yahoo-model" – altså et klikbart hierarki. Men en vis tøven for om det lod sig gøre på nationalt plan betød, at der ikke fra start blev tænkt i disse baner.

I 2003 fik DBC henvendelser fra flere biblioteker/bibliotekssystemer angående emnehierarkier. Emnehierarkier til slutbrugere var på vej i flere systemer og deraf fulgte naturligt nok overvejelser, om der virkelig skulle laves flere forskellige almene emnehierarkier i Danmark.

Fra medio 2003 til medio 2004 har der været drøftelser og høringer fra Biblioteksstyrelsen om relevansen af et "nationalt emnehierarki" til slutbrugere af bibliotek.dk – og til brug i lokale kataloger og evt. netguides. Her gengives nogle udpluk fra høringsmaterialet (link til det samlede materiale ses til slut i artiklen):

"Skabelsen af et brugervenligt nutidigt verbalt hierarki er en stor opgave. Vedligeholdelsen er også en større opgave. Det vil være venligt over for de brugere, der bruger flere biblioteker, at danske biblioteker bruger samme hierarki – og det vil være rationelt kun at udvikle det én gang."

Hvorfor skal vi have det? – er Google og klassifikation ikke nok?

Google-modellen

Søgemaskiner som Google og kataloger som bibliotek.dk er gode til de bevidste og præcise forespørgsler - det er sværere, når man ikke har en præcis søgeformulering.

Klassifikation

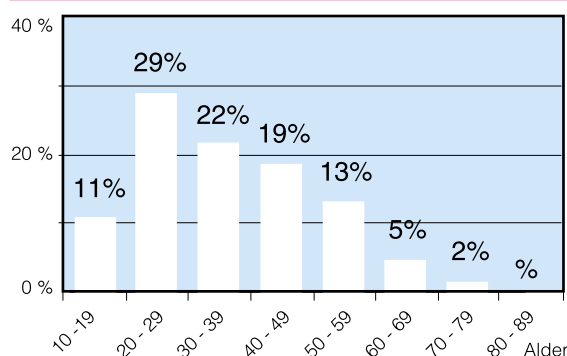
Klassifikation som fx dk5 giver en ordning af verdensbilledet og i princippet er det en sådan ordning, der er behov for. Men da klassifikation som oftest også danner baggrund for opstilling, vil der typisk være en stor træghed i udvikling af dette verdensbillede. Det kan aktuelt illustreres ved, at den verbaliserede udgave af dk5's øverste niveau ikke indeholder ord som fx: edb, internet, web, teknologi – eller sundhed, sygdom. Hvordan finder brugeren ud af, hvor han skal klikke – og hvor mange forgæves klik skal han igennem, inden han rammer rigtigt?

Klik på en hovedgruppe for at udfolde hierarkiet

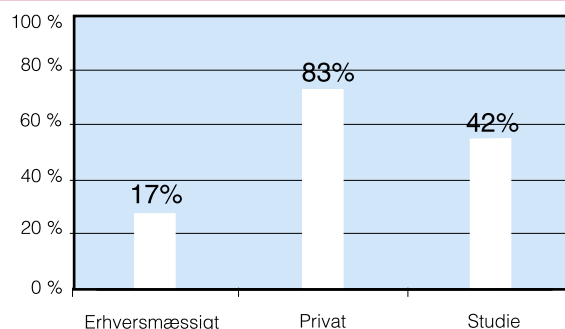
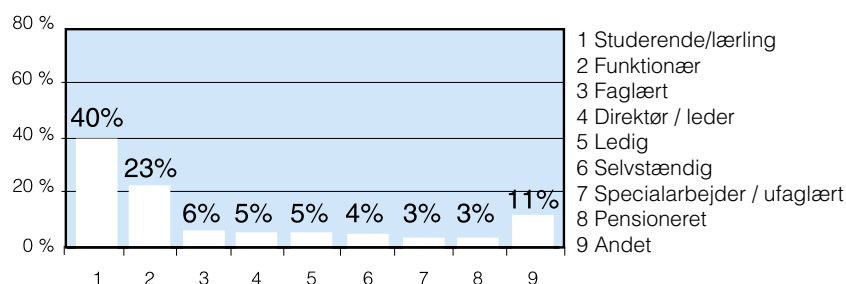
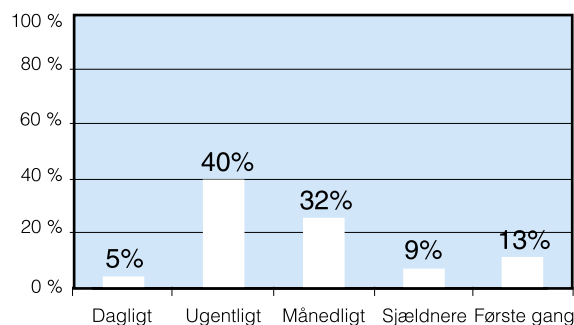
00-07 Værker af almindeligt og blandet indhold
10-19 Filosofi. Psykologi. Videnskab og forskning
20-29 Religion
30-39 Smfundsvidenskab. Pædagogik. Folkekultur
40-49 Geografi og rejser. Lokalhistorie i alm.
50-59 Naturvidenskab og matematik
60-69 Praktiske fag
70-79 Kunst. Teater. Film. MUsik. Spil. Idræt
80-89 Litteratur. Sprog
90-99 Historie

Emneguiden

Emneguiden med verbale, klikbare hierarkier kan hjælpe den bruger, som ikke har et præcist søgeord, eller den der har et alt for generelt begreb. Det kaldes populært Yahoo-modellen, men Yahoo har ændret sig en del. På fx www.dmoz.com kan ses et standardeksempel på et

Graf 1: Aldersfordeling**Graf 2: Ærinde**

- i hvilket ærinde besøgte du hjemmesiden?

**Graf 3: Beskæftigelse****Graf 4: Besøgsfrekvens**

universelt emnehierarki. Emnehierarkier kan ændres løbende, uden at opstillingen af materialerne ændres og de behøver ikke stile mod at dække ethvert emne.”

Formål med projektet Emnehierarki

At skabe et dansk verbalt emnehierarki til indgang primært til bibliografiske poster (sekundært også til andre typer data fx netguider) til den almindelige danske borger – brugere af bibliotek.dk, af emneguider, af lokale bibliotekssystemer. Det skal emnemæssigt være universelt. Der er ikke tale om et hierarki, som dækker ethvert emne, men et hierarki som kan hjælpe en del af de brugere, der har vanskeligt ved at finde det de leder efter ved menu- eller kommandosøgning. Det udarbejdes i første omgang til bibliotek.dk, men tilrettelægges så import til lokale kataloger og andre egne fælles-biblioteksprodukter er mulige.

Hvem kan bruge det og hvordan?

Den målgruppe emnehierarkiet skabes til er bibliotek.dk's brugere. De er defineret som danske biblioteksbrugere fra 13 år og opefter. Studerende er formentlig den største enkelt-gruppe af brugere, men hierarkiet skal ikke først og fremmest tilgodese deres behov, som i hvert fald nogle gange er at finde konkrete titler. Emnehierarkiet skal tilgodese søgebehov, som ikke er så

præcise, at de nuværende søgemetoder kan bruges. Det handler altså IKKE om, at den studerende kan finde sine referencer via emnehierarkiet, men om at en bruger, som IKKE har en præcis titel at lede efter, får en alternativ søgemåde.

I hørings svarene var der to hovedtendenser:

- tilslutning til ideen (især fra folkebiblioteker)
- modstand eller tøven over for ideen (især fra forskningsbiblioteker)

Biblioteksstyrelsen besluttede efter anden høringsrunde og flere drøftelser, at projektet skulle gennemføres – men med understregning af at det ikke kunne eller skulle tilgodese alle behov, men primært den ”almindelige brugers” behov for at finde materiale om emner, hvor brugeren IKKE har en klar fornemmelse af emnets terminologi eller struktur. Eller helt elementært: Til brugeren der ikke kan stave til det de søger – eller huske hvad det nu er, det hedder.

Et par citater fra brugerkommentarer til bibliotek.dk:

„ville være rart med en struktureret emneindgang som supplement til søgemaskinen”

„maskinen har svært ved at finde frem til bøger, hvis den ikke lige får de søgeord, den helst vil have”

Emnehierarkiets formål er at tilgodese:

- det mudrede informationsbehov
- behovet for overblikslitteratur
- for at finde nogle få gode titler
- hjælp når man ikke kan stave
- mulighed for inspiration

Udviklingsplan for emnehierarki

I 2004 blev selve hierarkiet udviklet (ordhierarki med tilhørende søgestreng) – i 2005 udvikles brugergrænseflade, driftsrutiner til eksport af emnehierarki til bibliotekssystemer og andre interesserede, samt opdatering af emnehierarkiet.

Vi har måtte basere os på eksisterende data – der er ikke råd til ekstra indeksering. DBC/folkebiblioteksdata bliver kernen - fordi der er en vis ensartethed - fordi det er et overskueligt sted at starte – og fordi forskningsbibliotekerne har været tvivlende/afvisende over for anvendeligheden dér. Hierarkiet skal være universelt, men ikke abstrakt. Sprogbrugen skal være nutidig, men ikke poppet. Vi er endt med 13 hovedgrupper – og i alt omkring 2000 ”emner”.

De søgninger, der ligger bag emnehierarkiets ord, kombinerer klassifikation, emneord, titler og (lidt) fritekst. Der er hovedfokus på at finde danske, centrale materialer – ikke ”alt materiale om”.

Indeksørernes erfaringer, kan vi alle lære af:

- kender man 5 søgekoder, kan man lave gode emnesøgninger og finde den centrale danske litteratur: ok, dk, em, ke, df
- hyppigst er en kombination af en klassifikationskode og en emnekode (titel - sjældnere fritekst)

Nogle eksempler fra hierarkiet

Naturkatastrofer

ok=55? og (naturkatastrofe? eller katastrofe? eller stormflod? eller tørke? eller vulkanudbrud? eller jordskælv? eller tsunami? eller flodbølge?)

Spam og firewall

ok=(19.639 eller 19.625 eller 62.38) og internet? og ((spam* eller firewall) eller em=*sikkerhed)

Indretning – badeværelser

em=badeværelser og ?indretning og dk=64?

Afrikas selvstændighed

(em=afrika eller dk=(98.3 eller 98.4?)) og (selvstændighed eller uafhængighed eller afkolonisering)

Udviklingshjælp til Afrika

(afrika? og dk=(32.72 eller 33.26:2)) eller (dk=(98.3 eller 98.4?) og (ulandsbistand eller u-landsbistand))

Hvem er bibliotek.dk-brugerne?

Det er nok rimeligt med et kort overblik over, hvem der bruger bibliotek.dk. Målgruppen er defineret som borgere fra 13 år og opefter, men brugerne tegner sig ifølge den nyeste undersøgelse (oktober 2004) sådan her:

- 62% er under 40 år, en rimelig ung brugergruppe.
- 42% bruger sitet i studieøjemed
- 83% i privat (man kunne naturligvis angive mere end ét formål)
- 40% af brugerne er studerende.
- 45% kommer mindst en gang ugentligt.

Samtidig stiger brugen hele tiden, så selvom der altså er rigtig mange ”faste kunder”, er der også hele tiden nye.

Hvad kan vi så bruge den viden til? Ja, blandt andet en erkendelse af at de studerende er en vigtig brugergruppe, selvom den fra bibliotek.dk's start ikke var defineret som en særlig vigtig målgruppe. Målgruppen er stadig vældig bred og formålene mange. Nogle brugere kommer på bibliotek.dk vældig ofte – men der er også hele tiden mange helt nye brugere. Vi kan altså

ikke definere bestemte forudsætninger som givet.

Hvordan søger bibliotek.dk-brugerne?

En meget stor del af brugen ser ud til at gå efter helt konkrete titler. Selvom brugere er rigtig glade for ”Google-modellen” med én søgeboks, har vi ikke valgt den på bibliotek.dk. (Der findes en fritekstboks, men det er ikke det eneste på søgesiderne). Vi kan ikke rankere posterne, som Google med et vist held kan, med de ”mest relevante” først. Selvom PageRank også blev problematiseret på temadagen, er der næppe tvivl om at til mange søgeformål virker Googles ranking brugbar for mange brugere. Derfor forsøger vi at ”tvinge oplysninger ud af brugerne” ved menu-søgningsbokse. Og det virker! En hurtig logfilanalyse af en dag i december (godt 20.000 søgninger) giver dette resultat:

Fritekst	07%
Forfattersøgning (boks og kommando)	34%
Videresøgning (link) på forfatter	5%
Titelsøgning	20%
Emnesøgning (boks og kommando)	11%
Videresøgning (link og ”Litteratur om..”) på emne	15%
Resten (hver enkelt max 2 %) i alt	8%

Brugerne kan altså godt ”overtales” til at bruge ”præfabrikerede søgninger” og ”søgekoder”, hvis det blot serveres lidt enkelt – og virker. Jeg tror ikke at det holdt, hvis brugerne syntes at fritekst-søgninger gav dem et bedre resultat end at bruge bokse, link, titler der ligner osv. Denne hurtige analyse giver altså et vist belæg for, at de også vil bruge emnehierarkiet, når og hvis det er relevant for dem.

Hvordan skal emnehierarkiet præsenteres?

Det bliver næste opgave i forbindelse med ny brugergrænseflade til

bibliotek.dk. Det er vigtigt, at få præsenteret de to søgemåder – den nuværende og emnehierarkiet – på den rette måde, så man vælger det relevante i forhold til søgebehovet. De to nuværende implementører af et dansk emnehierarki (DDE-Libra og Integra) gør ikke rigtig noget for at forklare, hvornår man bruger hvad – det er heller ikke let at forklare brugerne, som erfaringsmæssigt jo ikke bruger meget mere end et splitsekund før de handler. Ar-

bejdet med ny brugergrænseflade kommer til at vare det meste af 2005.

Eksport af emnehierarkiet

Projektet omfatter også at emnehierarkiet skal kunne eksporteres – til biblioteker, bibliotekssystemer eller netguider. Der er lavet et xml-baseret eksportformat (i samarbejde med systemleverandørerne), og samtidig med at emnehierarkiet i bibliotek.dk sættes i søen (til efteråret), vil man kunne abonnere på emnehierarkiet (kun betaling af håndteringsomkostninger – og da det bliver en fil på DBCs posthus, bliver det en mindre sag).

Hvad så med de proffe?

I forbindelse med undersøgelse af bibliotek.dk-brugernes søgeadfærd, har jeg også kigget på hvordan der søges i DanBib via netpunkt. Det var faktisk så ”interessant”,

at jeg vender tilbage til dette emne i en senere artikel!

Høringsmateriale om emnehierarki på BS's hjemmeside: www.bs.dk

Brugerundersøgelse fra 2004 på BS's hjemmeside: www.bs.dk

